



**The University of Cape Town**

**Department of Statistical Sciences**

**Faculty of Science**

**Using state-space time series analysis on wetland bird species to  
formulate effective bioindicators in the Barberspan wetland**

By

**Gareth Edwards**

Supervisors

**Res Altwegg**

**Birgit Erni**

Minor Dissertation presented for the degree of M.Sc. Data Science  
in the

Department of Statistical Sciences

Faculty of Science

University of Cape Town

January 2022

## Abstract

An abstract goes here

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Method</b>	<b>4</b>
2.1	Study Area . . . . .	4
2.2	Exploratory Data Analysis . . . . .	4
2.3	State-Space Time Series Model . . . . .	5
2.4	Biodiversity indices . . . . .	7
<b>3</b>	<b>Results</b>	<b>9</b>
<b>4</b>	<b>Bibliography</b>	<b>15</b>

# 1 Introduction

With the ongoing concern about biodiversity loss, monitoring biodiversity at various levels of biological organisation is important (Pereira et al. 2013). At the organismal level, biodiversity is generally measured by three components: number of species, total number of individuals (abundance), and how the individuals are distributed across species (evenness). We lose biodiversity if abundance decreases, if species go extinct, if communities become dominated by a few common species, or a combination of these processes. To measure changes in biodiversity, we need indicators that are sensitive to changes in these components (S. T. Buckland et al. 2005).

Let  $n_i$  be the number of individuals of species  $i$  and  $n = \sum_{i=1}^S n_i$  the total number of individuals of  $S$  species in the community, so that  $i = 1, \dots, S$ . Evenness measures how uniform the species proportions  $p_i$  are, where  $p_i = n_i/n$ . Now consider surveying a single location and counting the individuals of a group of species over time. The counts  $n_{i,j}$  are the number of individuals encountered of species  $i$  in year  $j$ . Because population growth is a multiplicative process, changes in  $n_i$  over time are best measured as ratios  $n_{i,j+1}/n_{i,j}$ . Trends in the abundance of individual species can then be aggregated using the geometric mean (Stephen T. Buckland et al. 2011) or equivalently, the arithmetic mean of the changes in log abundances  $G_j = \exp\left(\frac{1}{S} \sum_{i=1}^S \log \frac{n_{i,j+1}}{n_{i,j}}\right)$ . The well-known Living Planet Index, for example, is based on the geometric mean (Loh et al. 2005; McRae, Deinet, and Freeman 2017).

The LPI is a population trend aggregation method. This means that the global LPI measure reflects an aggregated trend of various vertebrate species on Earth over time (McRae, Deinet, and Freeman 2017). However, biodiversity does not have just one definition, but instead encapsulates a broader concept where there are multiple metrics needed to capture different aspects of biodiversity (Gaston and Spicer 2013). Biodiversity measures are commonly broken up into two basic components. That is, how many different species are present in an ecosystem (Magurran 2003) and how many individuals per species are present in an ecosystem (evenness) (S. T. Buckland et al. 2005). Based on these components of biodiversity, a community with more species tends to have higher biodiversity than a community with fewer species and a community that displays high levels of evenness is regarded more diverse than a community with the same number of species but that displays high levels of unevenness. Regarding the concept of evenness, if two communities have the same number of species, the one where each species is equally abundant (high degree of evenness) is more diverse than one where there are a few common species and the rest are rare (high degree of unevenness). The exponentiated Shannon's index and the Simpson's index are the two most common indices used to measure effective number of species in an ecosystem, and ecosystem evenness, respectively (Hill 1973).

The above mentioned indices can be beneficial in monitoring ecosystem health. Studies have shown that monitoring the change in diversity of bird species in wetland ecosystems could potentially be beneficial in determining wetland health (Amat and Green 2010; Gregory and Strien 2010; Sekercioglu 2006). For example, Birds have been used to indicate the eutrophication levels in wetlands in the Mar Menor lagoon of south-eastern Spain (Amat and Green 2010). It was discovered that the Great Crested Grebe abundance increased as the eutrophication levels in the lagoon increased. Another example from Southern Spain shows the use of Red-Knobbed Coot to indicate the siltation rate in the water and soil in wetlands (Amat and Green 2010). The Red-Knobbed Coot experienced a stark decline in abundance in the 20th century in Southern Spain. This decrease in abundance is attributed to the increased siltation rates in the water and soil in the wetlands in Southern Spain, caused by invasive agricultural practices.

Birds are valuable members of any ecosystem and their count data follows exactly what is needed for a good bio-indicator as effective bio-indicators are quantitative, easy to collect, simple to interpret, reflect causes of change in an ecosystem and indicate general patterns of taxa in the ecosystem (Sekercioglu 2006; Gregory et al. 2003). In this situation we refer to bio-indicators as species whose fluctuation in abundance closely reflects the state of the ecosystem they find themselves in. The pitfall to using birds as bio-indicators is that the count process is often prone to error, either under counting or over counting (Auger-Méthé et al. 2021). When using erroneous counts to calculate environmental indices we see the errors carrying forward through the calculations, and inevitably, negatively affecting the result (Kéry and Schaub 2011).

Observation error is a hindrance, but can be accounted for when using the correct model. One such model is state-space models (SSMs). SSMs are hierarchical models that are able to decompose observed time series data into a process variation and an observation error component (Kéry and Schaub 2011). When applying a State-space time series model within a bayesian context, one is able to output a posterior distribution of counts for each time period (Auger-Méthé et al. 2021). One can then use these posterior counts to calculate various biodiversity indices such as the LPI, exponentiated Shannon's index and the Simpson's index. An added benefit is that we can produce credible intervals for these index values, a quality that is not commonly incorporated in such calculations.

The aim of this paper is to present a method of calculating and presenting biodiversity indices by first fitting a State-space time series model to the count data and then using the posterior output to calculate the biodiversity index. In this paper the indices calculated are a modified LPI, an exponentiated Shannon's index and a Simpson's index with their respective credible intervals. The

data being used for this report is from the Coordinated Waterbird Counts (CWAC) project. CWAC is a citizen scientist initiative that contains water bird counts of over 200 different bird species from approximately 400 different wetland locations in South Africa. Some of these counts go as far back as the 1970s. This report will focus on the bird counts from the Barberspan wetland in North West Province of South Africa.

## 2 Method

### 2.1 Study Area

The bird counts used in this study come from the Barberspan wetland which is the pilot site for this study. The Barberspan wetland is situated in North West province, South Africa. It is a protected wetland site known as Barberspan Bird Sanctuary that covers a 3200ha area. The wetland is centered on a shallow lake which covers approximately 257ha to 2000ha depending on rain fall (Remisiewicz and Avni 2011).

RAMSAR reports that the Barberspan wetland regularly supports approximately 20000 waterbirds with 365 different bird species (Rothmann J 2015). Barberspan is one of the first RAMSAR sites in South Africa and the only RAMSAR site in North West province.

Barberspan wetland is home to thousands of Little Stints and Curlew Sandpipers. It's also a popular location for intra-African nomadic waders such as Kittlitz's Plover and Crowned Lapwing. The bird species with the highest abundance in Barberspan wetland is the Red-Knobbed Coot, Egyptian Goose, Yellow Billed Duck, Southern Pochard and the South African Shelduck (Remisiewicz and Avni 2011). Appendix A contains further information regarding the waterbirds found in Barberspan wetland that was used in this study. The Barberspan wetland is rich in bird data and thus is an ideal wetland to use as a pilot site for this study.

### 2.2 Exploratory Data Analysis

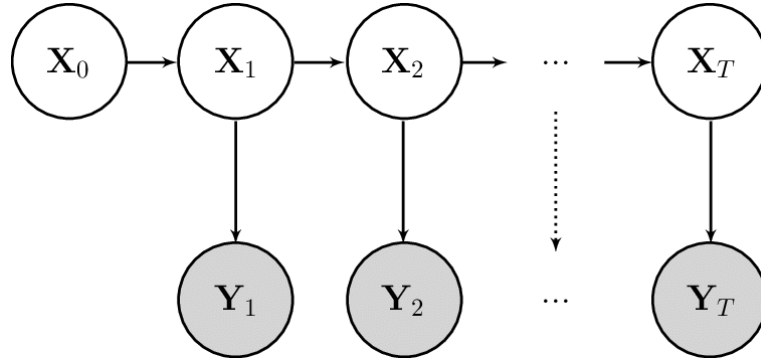
The data used in this paper is from the Coordinated Waterbird Counts (CWAC) initiative. The counts are conducted twice a year, once in January/February (austral summer) and once in June/July (austral winter). The counting procedure is standardised and follows a rigorous set of instructions that counters are to follow. The counting protocol is such that all counts take place at approximately the same time each day, when the sun is either behind or beside the counter. More details about the counting process can be found at the following link (<https://cwac.birdmap.africa/>).

The CWAC dataset contains bird counts from 1993 to 2013 for the Barberspan wetland. Missing counts were inserted into the dataset as missing values that will subsequently be predicted by the state-space time series model. Only birds with 15 or more non-zero counts were used in the analysis as the variance parameters for the state-space time series model didn't converge when applied to bird counts with fewer than 15 non-zero counts.



### 2.3 State-Space Time Series Model

A state-space time series model (SSM) was fitted to the bird count data. State-space models have become increasingly popular tools in modelling population dynamics (Auger-Méthé et al. 2021). The key benefit of a state-space model is its ability to separate observation error and latent error. By removing the observation error from the latent error, one is able to predict population trends more accurately (Kéry and Schaub 2011). SSMs are an ideal tool to use for time series analysis for the CWAC dataset, as bird counts are often prone to high levels of observation error and follow a markovian process, where the value at time  $t$  is dependent on the value at time  $t-1$  (see figure 1 for illustration of markovian chain).



**Figure 1:** A visual representation of a state space model where  $X_t$  represents the unobserved population level changes. SSMs use the observed population counts,  $Y_t$ , to estimate the population level changes.

In this study the CWAC bird counts are modeled using an existing state-space time series model that was created by Barshep et al. (2017) and applied to bird count data in the CWAC dataset. Barshep et al. (2017) developed a bivariate time series model that treats summer and winter counts as separate variables, with the focus being on the summer population. The winter counts are modeled as a proportion of the summer counts as some birds from Barberspan migrate north during the southern winter.

SSMs model the state process and observation process separately such that state error and observation error can be separated for more accurate modelling of the count data. The mathematical illustration of the bivariate model is displayed below:

*observation process*

$$\ln(\text{summer count}) : s_t = \mu_{t-1} + \alpha_t, \quad \alpha_t \sim N(0, \sigma_\alpha^2) \quad (1)$$

$$\ln(\text{winter count}) : w_t = \mu_t + \lambda_t + e_t, \quad e_t \sim N(0, \sigma_e^2) \quad (2)$$

*state process*

$$\ln(\text{summer population level}) : \mu_t = \mu_{t-1} + \beta_{t-1} + w_t, \quad w_t \sim N(0, \sigma_w^2) \quad (3)$$

$$\ln(\text{population change}) : \beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim N(0, \sigma_\zeta^2) \quad (4)$$

$$\ln(\text{summer to winter ratio}) : \lambda_t = \lambda_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (5)$$

where,  $s_t$  is the summer count at time  $t$  and  $w_t$  is the winter count at time  $t$ .  $\mu_t$  is the latent summer population,  $\beta_t$  is the rate of change at time  $t$  and  $\lambda$  is the winter to summer ratio. The winter to summer ratio refers to the difference in the bird population from summer to winter. The palearctic migrants usually migrate north for winter, and leave their young behind, thus leaving a smaller proportion of birds in the South African wetlands in winter compared to summer.  $\alpha_t$ ,  $e_t$ ,  $w_t$ ,  $\zeta_t$  and  $\epsilon_t$  are all normally distributed, serially independent and independent of each other with variance  $\sigma_\alpha^2$ ,  $\sigma_e^2$ ,  $\sigma_w^2$ ,  $\sigma_\zeta^2$  and  $\sigma_\epsilon^2$  respectively.

The time series model presented above follows a Markovian process as the value at time  $t$  is dependent on the value at time  $t-1$ . The aim of this analysis is to sample from the probability distributions that represent the summer and winter counts. The current distribution for these summer and winter counts are unknown, but by applying Monte Carlo simulations from the summer and winter count distributions we can generate values that are from an approximation of the true count distribution (the stationary distribution). This process is known as a Monte Carlo Markov Chain (MCMC) approach (Brooks et al. 2011). The simulated values are known as the posterior distribution and is later used to calculate bio indices with credible intervals.

In this analysis the MCMC algorithm used is the Gibbs Sampler, which is applied using the JAGS package in R (Plummer 2003; **R?**; **jagsui?**). The JAGS software implements Bayesian inference based on Gibbs sampling by using a MCMC approach (Coro 2017). The hyperparameters available in the JAGS program are number of chains (n.chains), number of iterations (n.iter) and burn-in length (n.burnin). The number of chains refers to the number of markov chains created. These are the number

of chains that run in parallel which hopefully converge to form the stationery distribution. A trace plot can also be plotted to show if these chains converge or not for a given summer count value. If all chains converge around the same approximation then we can be satisfied that the model has reached convergence. The number of chains was set to three in this analysis. The hyperparameters for number of iterations and “burn-in” for the Gibbs Sampler was 10 000 and 5000 respectively. The number of iterations refers to the number of times the Gibbs Sampler samples from the stationery distribution that represents the summer or winter counts. The “burn-in” hyperparameter refers to the first few iterations that are to be discarded given that these first few iterations are usually less probable outcomes, but as the model runs for a longer period of iterations, the outcomes tend toward a higher probability region and thus are more likely outcomes (Kéry and Schaub 2011). This generates an output of 15 000 count values for each bird for each year with the first 5000 iterations being the “burn-in” iterations.

## 2.4 Biodiversity indices

The 10 000 approximated count values for each bird count per year was used in calculating the modified LPI, exponentiated Shannon index and the Simpson index. The benefit of using the whole posterior sample of bird counts is so that a credible interval could be calculated for each index.

A popular population abundance index is the Living Planet Index (LPI) that is designed to track changes in population sizes over time. The LPI is used as an aggregated abundance index. It calculates the percentage change from a reference year (1970) to all subsequent years. This is done by averaging the rate of change (from reference year to current year) across all species of a species group (McRae, Deinet, and Freeman 2017).

A modified LPI approach was used to form an abundance index in this study. The modified LPI used calculates percentage change from each year to the next for each bird type in Barberspan and then calculates the average for that year. This is unlike the standard LPI approach that calculates the rate of change of species for each year based on the reference year (1970).

The modified LPI calculations are displayed as follows:

$$\overline{d_t} = \frac{1}{n_t} \sum_{i=1}^{n_t} d_{it} \quad (6)$$

where  $\overline{d_t}$  is the average annual trend if there were multiple population time series of the same species.  $n_t$  refers to the number of populations.  $d_t$  is the annual rate of change for a population at time  $t$ , and

is represented as follows:

$$d_t = \log_{10} \left( \frac{N_t}{N_{t-1}} \right) \quad (7)$$

where  $N_t$  refers to the population size during year  $t$ .

The modified LPI calculations were applied to the summer population posterior values, for each year, estimated by the state-space model. The mean modified LPI was taken for each year and credible intervals were formed around that mean by calculating the 2.5th and 97.5th percentiles.

The exponentiated Shannon index and the Simpson index was also used to calculate further aspects of biodiversity in Barberspan. Namely the evenness and effective number of bird species. This is calculated using the Simpson's index and exponentiated Shannon's index respectively (Nagendra 2002). The calculations for the Shannon's index and the Simpson's index is displayed as follows:

*Shannon's index*

$$x = - \sum_{i=1}^S p_i \ln(p_i) \quad (8)$$

where  $x$  is the index,  $p_i$  is the proportion of species  $i$  in the sample and  $S$  is the number of species in the sample.

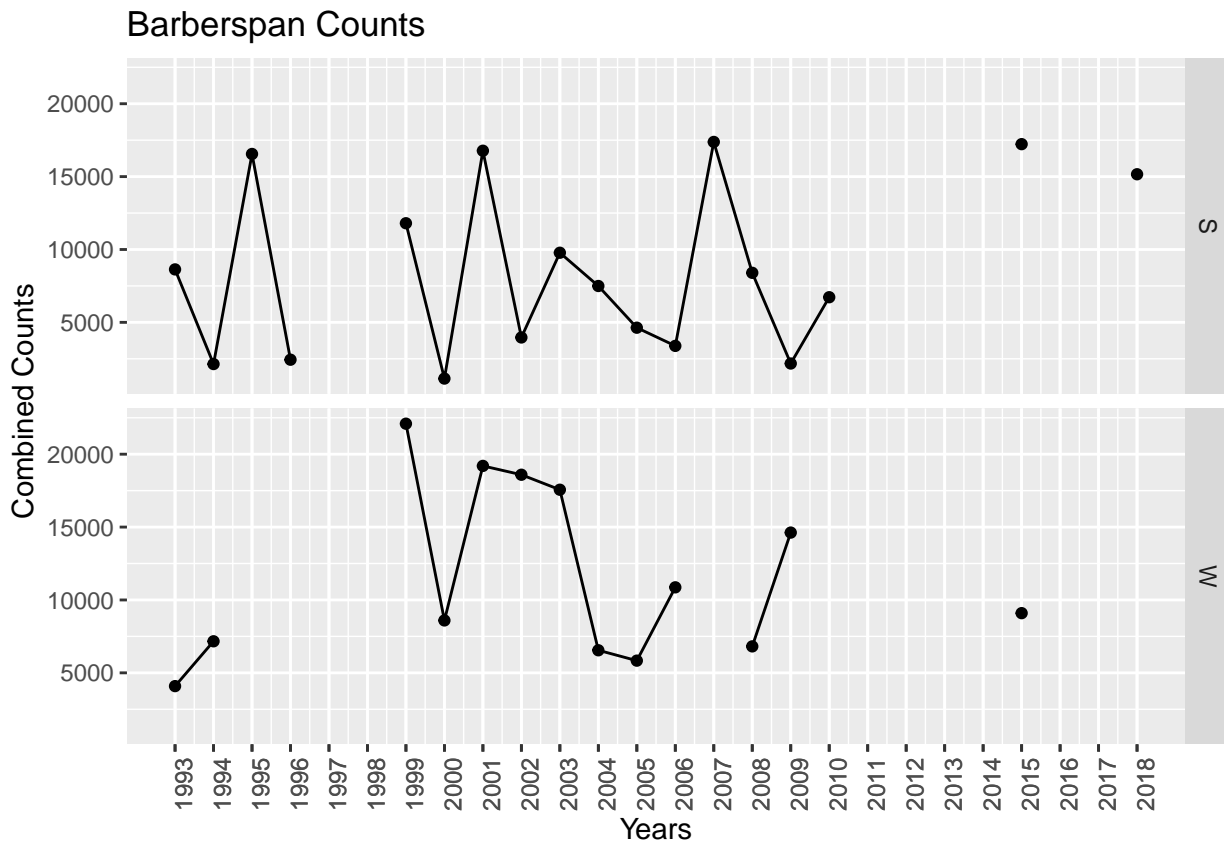
*Simpsons index*

$$x = \sum_{i=1}^S p_i^2 \quad (9)$$

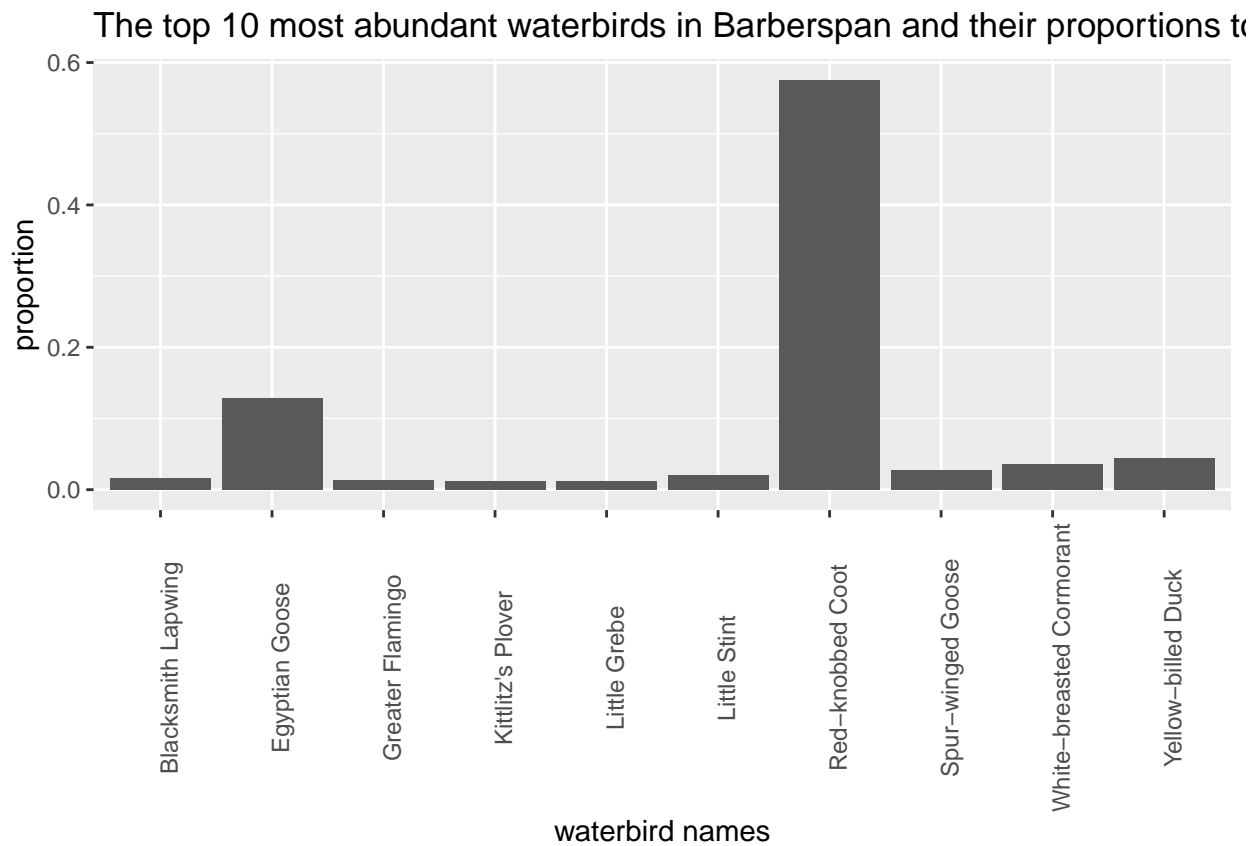
The Shannon and Simpson index were applied to the posterior values generated from the Gibbs sampler in similar manner as the modified LPI calculations. An exponentiated Shannon and Simpson index was calculated per year, using all 10 000 posterior count values for each bird species for that year. Thus creating 10 000 index values for each year. The mean index value was taken for both Shannon and Simpson values and the 2.5th and 97.5th percentiles were calculated to form the credible intervals.

### 3 Results

The combined bird counts of Barberspan from 1993 to 2018 are plotted in figure (x). The plot is broken up into summer and winter counts due to the biannual data capturing process that the CWAC protocols require. The plot shows fairly random variations of bird counts over the years, for both summer and winter counts, with multiple missing values in the dataset. Applying the state-space time series model to this data will approximate these missing values based on the previous existing values.

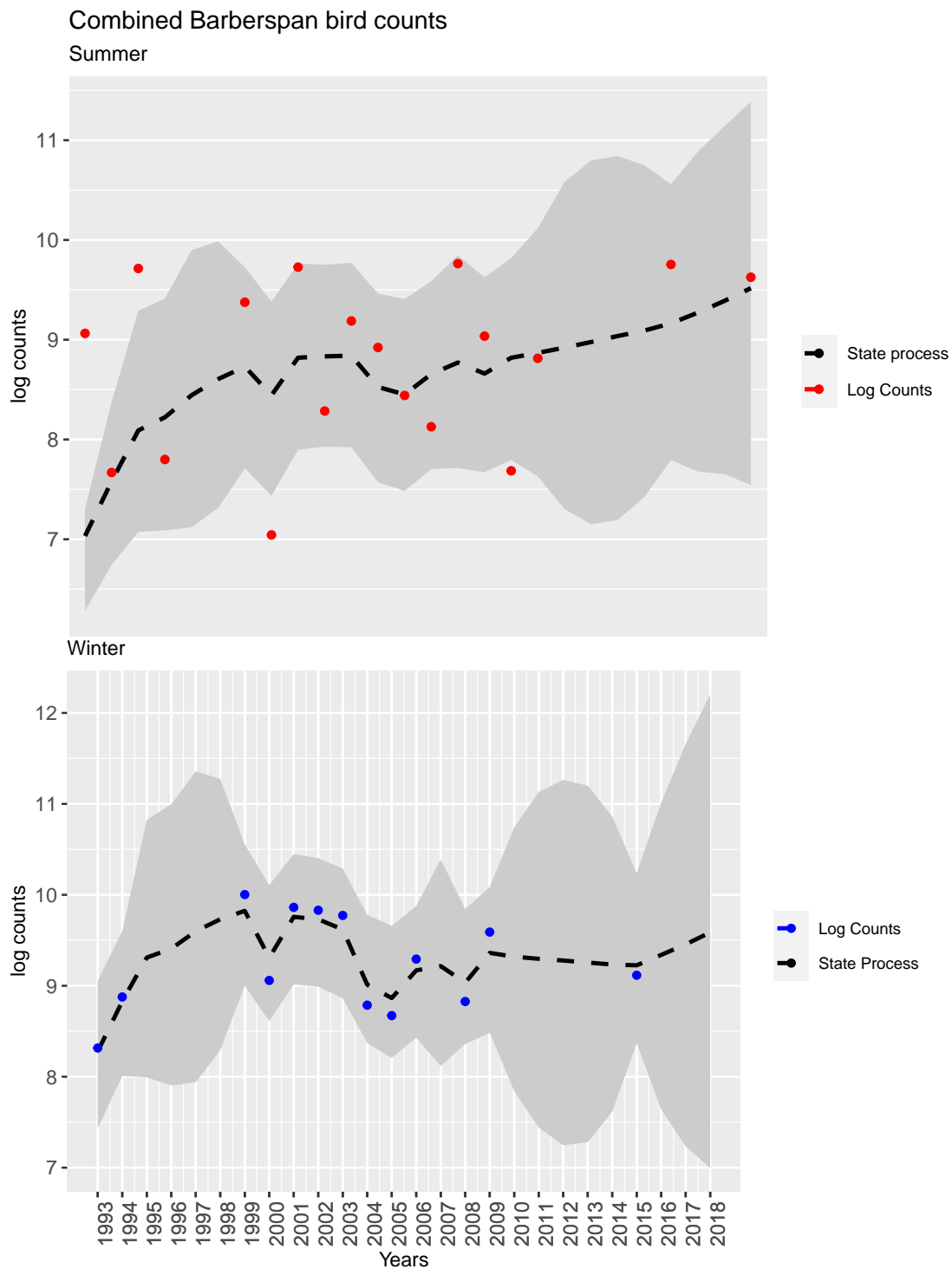


SHOW BAR PLOT OF 10 MOST ABUNDANT BIRDS IN BARBERSPAN



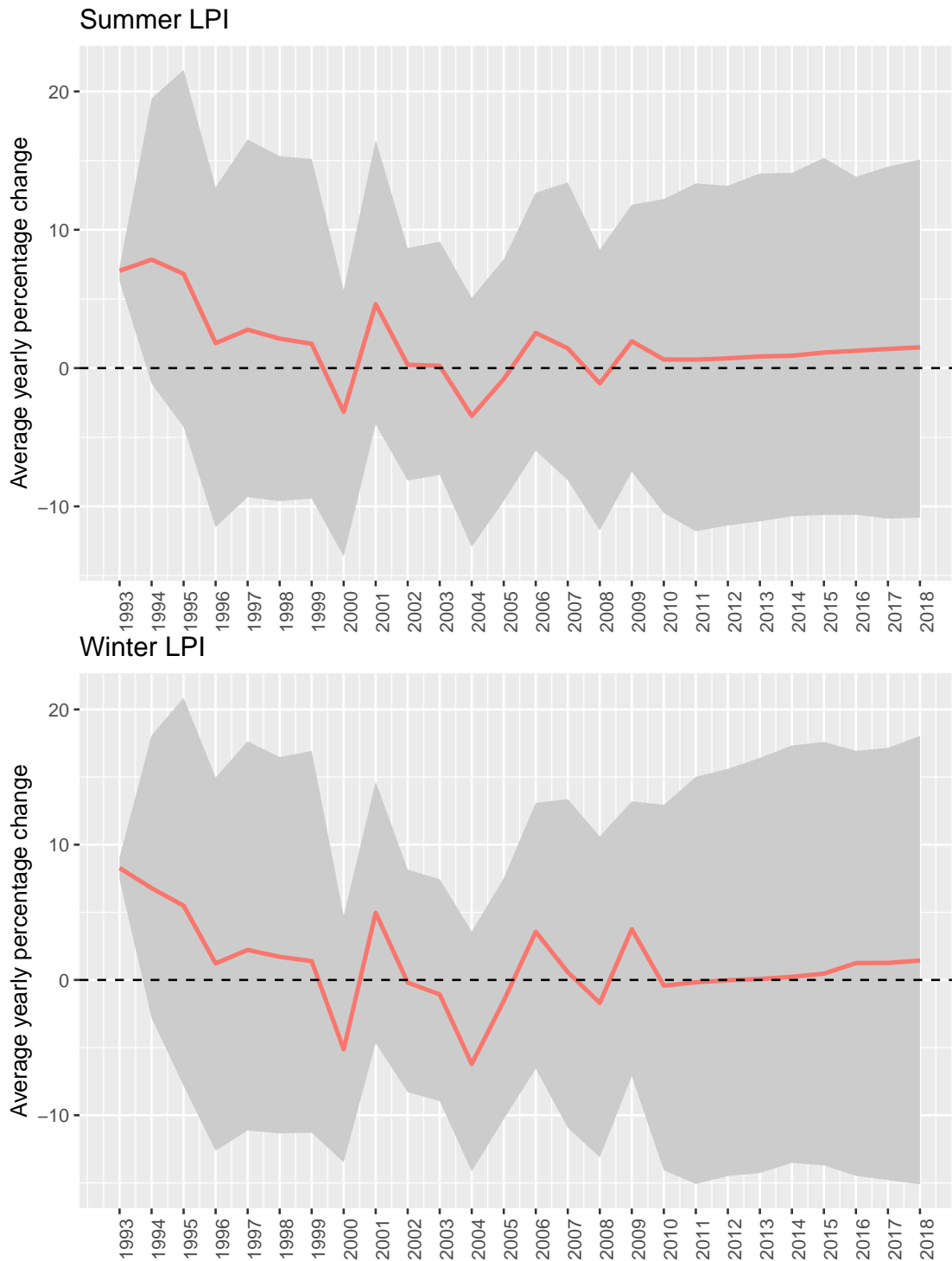
A state-space time series model was fit to the combined bird counts of the Barberspan wetland using the JAGS (Just Another Gibbs Sampler) package. The model was fit on the summer and winter counts separately. The output is posterior distributions for summer and winter counts for each year.

The state-space time series model output is displayed in figure (x). The diagnostic output of the state-space time series model is presented in Appendix (B) where trace plots of each MCMC chain is visualised to illustrate the convergence in count values for each year.



From the 10 000 yearly count values generated from the Gibbs Sampler, we are able to calculate various bioindices. We demonstrate the benefits of using posterior output to calculate bioindices by calculating the LPI, exponentiated Shannon index and the Simpson index using the 10 000 yearly posterior values.

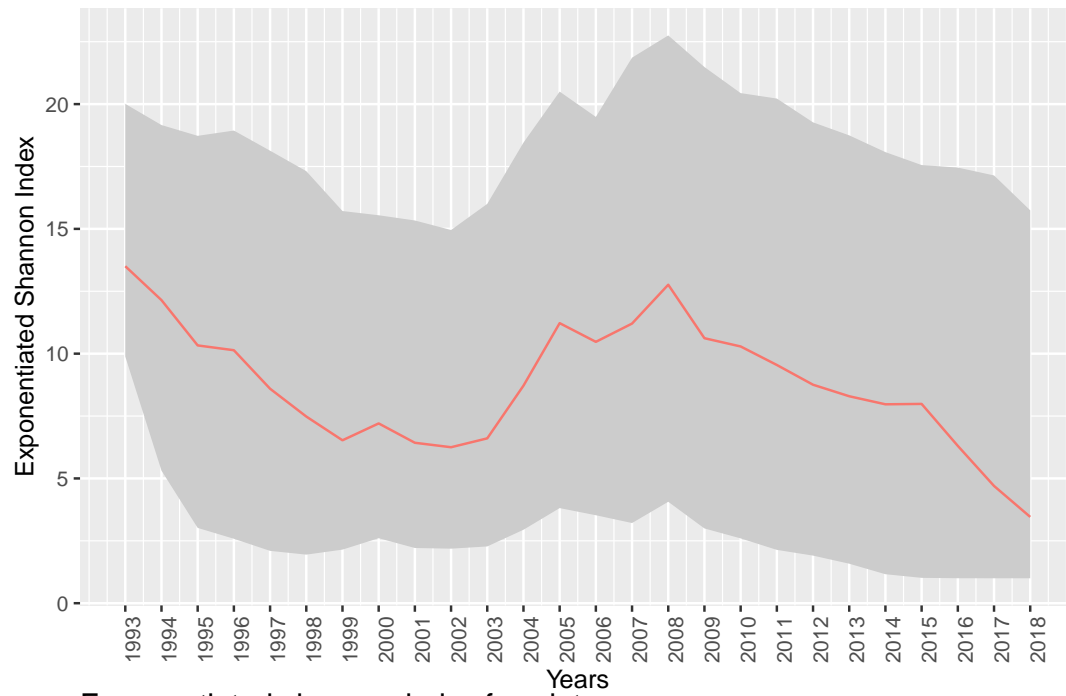
An LPI is calculated for each year and the mean, 2.5th and 97.5th percentile is generated from the 10 000 LPI values generated for each year. The output of this LPI calculations is shown in figure (x).



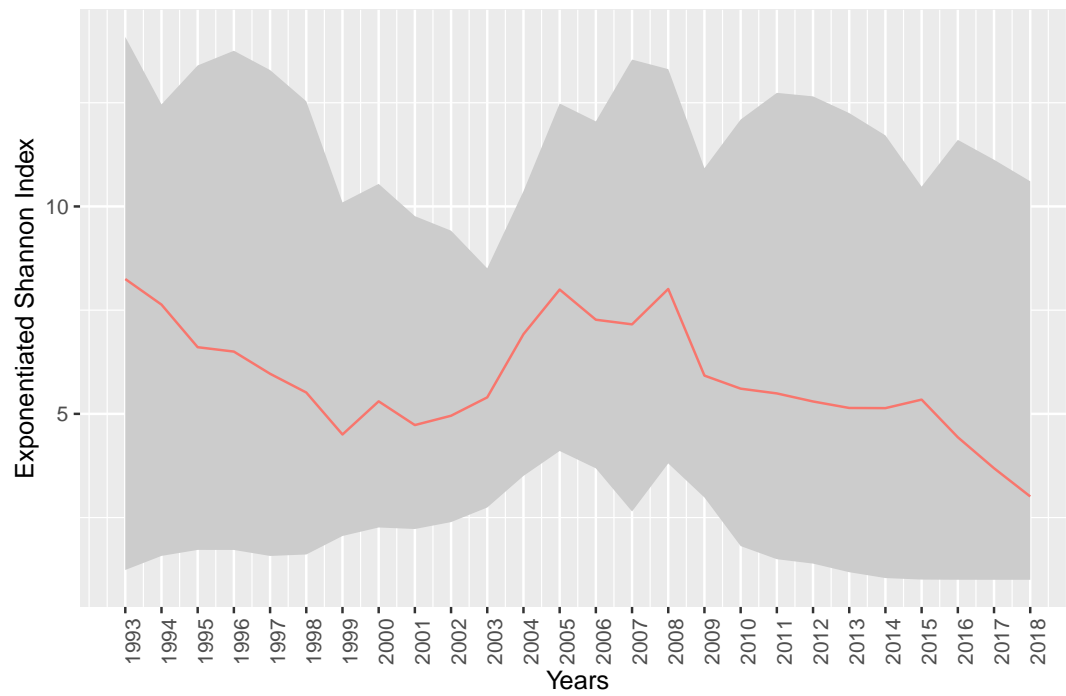
The exponentiated Shannon index was also calculated using the posterior output. An exponentiated Shannon index was calculated for each year. This resulted in 10 000 exponentiated Shannon index values for each year. The mean, 2.5th and 97.5th percentile was then calculated from the 10 000 index values. The same approach was conducted when calculating the Simpson index for each year. The outputs of the exponentiated Shannon and Simpson index calculations are displayed in figure (x) and (x2) respectively.



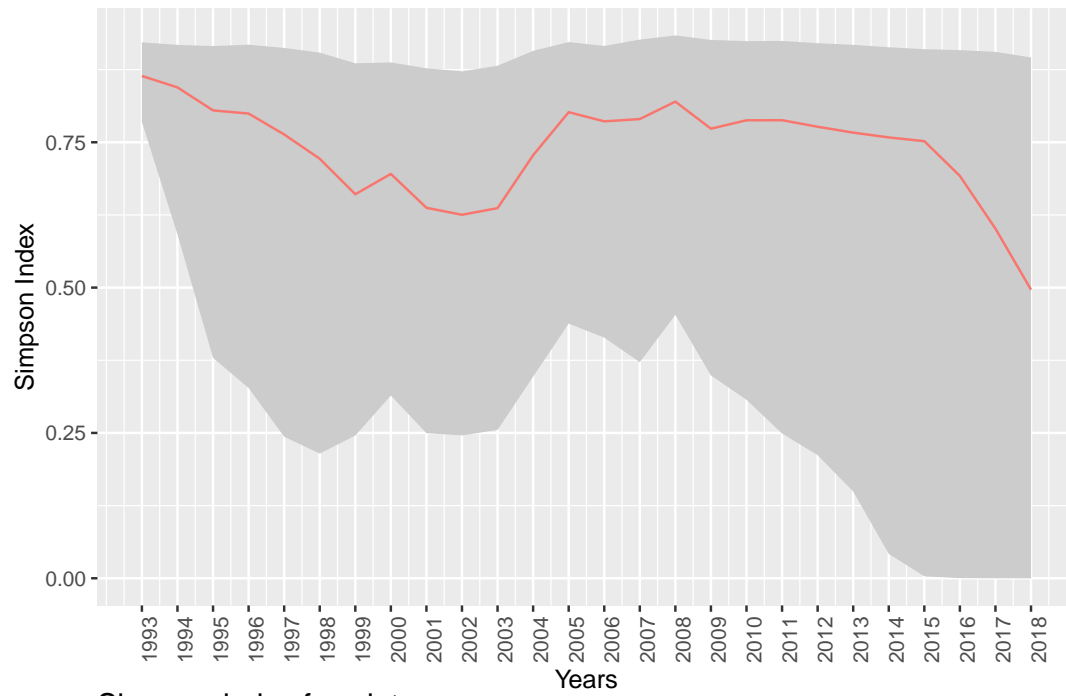
Exponentiated shannon index for summer



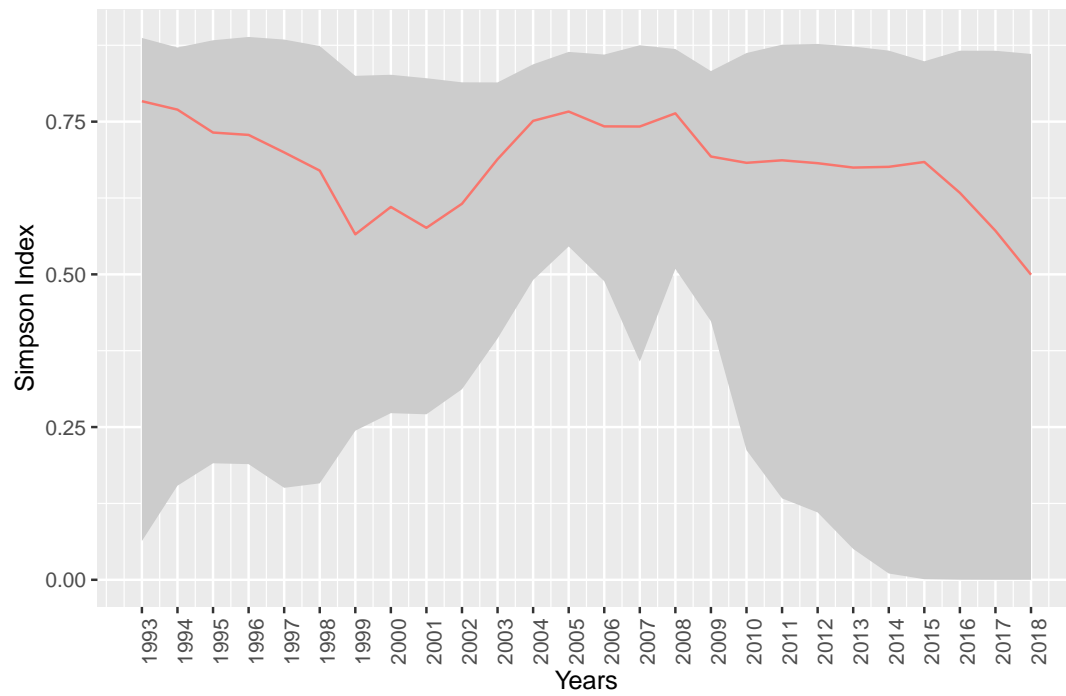
Exponentiated shannon index for winter



Simpson index for summer



Simpson index for winter



## 4 Bibliography

- Amat, J. A., and A. J. Green. 2010. “Waterbirds as Bioindicators of Environmental Conditions.” In *Conservation Monitoring in Freshwater Habitats: A Practical Guide and Case Studies*, edited by C. Hurford, M. Schneider, and I. Cowx, 45–52. Dordrecht: Springer Netherlands.  
[https://doi.org/10.1007/978-1-4020-9278-7\\_5](https://doi.org/10.1007/978-1-4020-9278-7_5).
- Auger-Méthé, M., K. Newman, D. Cole, F. Empacher, R. Gryba, A. A. King, V. Leos-Barajas, et al. 2021. “A Guide to State-Space Modeling of Ecological Time Series.” *Ecological Monographs* 91 (4).  
<https://doi.org/e01470>.
- Barshep, Y., B. Erni, L. Underhill, and R. Altwegg. 2017. “Identifying Ecological and Life-History Drivers of Population Dynamics of Wetland Birds in South Africa.” *Global Ecology and Conservation* 12: 96–107.
- Brooks, Steve, Andrew Gelman, Galin Jones, and Xiao-Li Meng. 2011. *Handbook of Markov Chain Monte Carlo*. CRC Press.
- Buckland, S. T., A. E. Magurran, R. E. Green, and R. M. Fewster. 2005. “Monitoring Change in Biodiversity Through Composite Indices.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1454): 243–54. <https://doi.org/10.1098/rstb.2004.1589>.
- Buckland, Stephen T, Angelika C Studeny, Anne E Magurran, Janine B Illian, and Stuart E Newson. 2011. “The Geometric Mean of Relative Abundance Indices: A Biodiversity Measure with a Difference.” *Ecosphere* 2 (9): 1–15.
- “Convention on Biological Diversity.” 2020. *Aichi Biodiversity Targets*.  
<https://www.cbd.int/sp/targets/>.
- Coro, G. 2017. “Gibbs Sampling with JAGS: Behind the Scenes.” [https://www.researchgate.net/publication/313905185\\_Gibbs\\_Sampling\\_with\\_JAGS\\_Behind\\_the\\_Scenes](https://www.researchgate.net/publication/313905185_Gibbs_Sampling_with_JAGS_Behind_the_Scenes).
- FitzPatrick Institute of African Ornithology. 2020. “CWAC Coordinated Waterbird Counts.”  
<https://cwac.birdmap.africa/>.
- Gaston, Kevin J., and John I. Spicer. 2013. *Biodiversity: An Introduction*. John Wiley & Sons.
- Gregory, R. D., D. Noble, R. Field, J. Marchant, M. Raven, and D. W. Gibbons. 2003. “Using Birds as Indicators of Biodiversity.” *Ornis Hungarica* 12 (13): 11–24.
- Gregory, R. D., and A. van Strien. 2010. “Wild Bird Indicators: Using Composite Population Trends of Birds as Measures of Environmental Health.” *Ornithological Science* 9 (1): 3–22.  
<https://doi.org/10.2326/osj.9.3>.
- Hill, M. O. 1973. “Diversity and Evenness: A Unifying Notation and Its Consequences.” *Ecology* 54 (2): 427–32. <https://doi.org/10.2307/1934352>.

- Kéry, Marc, and Michael Schaub. 2011. *Bayesian Population Analysis Using WinBUGS: A Hierarchical Perspective*. Academic Press.
- Loh, J., R. E. Green, T. Ricketts, J. Lamoreux, M. Jenkins, V. Kapos, and J. Randers. 2005. “The Living Planet Index: Using Species Population Time Series to Track Trends in Biodiversity.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 360 (1454): 289–95. <https://doi.org/10.1098/rstb.2004.1584>.
- Magurran, A. E. 2003. *Measuring Biological Diversity*. John Wiley & Sons.
- McRae, L., S. Deinet, and R. Freeman. 2017. “The Diversity-Weighted Living Planet Index: Controlling for Taxonomic Bias in a Global Biodiversity Indicator.” *PLoS One* 12 (1): e0169156. <https://doi.org/10.1371/journal.pone.0169156>.
- Nagendra, Harini. 2002. “Opposite Trends in Response for the Shannon and Simpson Indices of Landscape Diversity.” *Applied Geography* 22 (2): 175–86. [https://doi.org/10.1016/S0143-6228\(02\)00002-4](https://doi.org/10.1016/S0143-6228(02)00002-4).
- Pereira, Henrique Miguel, Simon Ferrier, Michele Walters, Gary N Geller, Rob HG Jongman, Robert J Scholes, Michael William Bruford, et al. 2013. “Essential Biodiversity Variables.” *Science* 339 (6117): 277–78.
- Plummer, M. 2003. “JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling.” *3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria* 124.
- Remisiewicz, M., and J. Avni. 2011. “Status of Migrant and Resident Waders, and Moulting Strategies of Migrant Waders Using African Inland Wetland Habitats, at Barberspan Bird Sanctuary in South Africa: Reports from BOU-Funded Projects.” *Ibis* 153 (2): 433–35. [https://doi.org/10.1111/j.1474-919X.2011.01112\\_1.x](https://doi.org/10.1111/j.1474-919X.2011.01112_1.x).
- Rothmann J. 2015. “Barberspan.” South {Africa}'s {Ramsar} {Sites}. *Barberspan*. <http://www.saramsar.com/2015/06/barberspan.html>.
- Sekercioglu, Cagan H. 2006. “Increasing Awareness of Avian Ecological Function.” *Trends in Ecology & Evolution* 21 (8): 464–71. <https://doi.org/10.1016/j.tree.2006.05.007>.